

Q.630.7
IL6C
no. 101
cop. 5
UNIVERSITY OF ILLINOIS
AGRICULTURE LIBRARY

CIRCULATING COPY
AGRICULTURE LIBRARY

UNIVERSITY OF ILLINOIS

Agricultural Experiment Station

URBANA, JANUARY, 1906

CIRCULAR NO. 101

METHODS OF TESTING VARIABILITY IN CORN

BY EUGENE DAVENPORT, PROFESSOR OF THREMMATOLOGY

The corn breeder needs some means of accurately testing the *range* and *extent* of variability and of knowing whether the variety he is breeding is approaching a common type as the years go by, or whether in spite of selection it tends strongly to wander; and if so, in what direction this tendency expresses itself. For example, ears of corn grow of different lengths, from the shortest nubbin to the longest ear. The number of rows is exceedingly variable, as are the diameter, circumference, and the weight of the ear, and neither the eye nor the memory is an accurate judge of either facts or tendencies.

The breeder selects according to the score card, somewhat arbitrarily. He does this without knowing whether, for example, the length of ear selected is above or below the racial type of corn; indeed there is among farmers no well recognized method for finding the racial differences between varieties. We can see at a glance that some varieties have longer ears than others, that some are larger than others, and that some are clearly more variable. But how much more? and do they remain constant year after year? or do they tend to deviate? and if they deviate, do they move together or do they not?

All these are important questions to the breeder for he needs to know everything possible about his corn. Recently developed methods in the study of variation make it possible to answer these questions. It is the purpose of this circular to explain these methods and show how variability may be accurately tested.

For example, it was desired to compare two varieties of corn. Two hundred ears of each were accurately measured as to length.¹ Of the first variety two ears were four inches long; four were five inches long; seven were six inches long; twelve were seven inches long; twenty-two were eight inches long; forty were nine inches long; sixty-five were ten inches long; thirty-four were eleven inches long; eleven were twelve inches long; and three were thirteen inches long.

Of the second variety, three ears were six inches long; eight were seven inches long; twenty-nine were eight inches long; ninety-four were nine inches long; forty-three were ten inches long; eighteen were eleven inches long, and five were twelve inches long.

How now do these varieties compare as to variability? To facilitate comparison and make possible the answer to this question we arrange the measurements as follows, in what is known as the "array":

Length in inches.....	4	5	6	7	8	9	10	11	12	13
Frequency, 1st var....	2	4	7	12	22	40	65	34	11	3—total...200
Frequency, 2nd var....	0	0	3	8	29	94	43	18	5	0—total...200

In this array the first line of figures represents the length of ears taken in inches, disregarding fractions. (In extreme accuracy we would use half inch differences, but experience shows that smaller divisions are unnecessary.) Under each length is placed the number of ears of that particular measurement in each variety. These numbers are known as frequencies. That is to say, for example, there are 22 ears of the first variety, and 29 of the second variety eight inches in length. Thus are the variations in the length of two varieties spread before the eye, the frequency distribution representing the variability.

Thus the numbers 2, 4, 7, 12, 22, 40, 65, 34, 11, 3, represent the range of variation in the first variety, and the corresponding numbers, 3, 8, 29, 94, 43, 18, 5, for the second variety.

The first noticeable point about these arrays is that the numbers are comparatively small at the extremes, rising to the largest values somewhere near the middle.

¹ A larger number, 400 or 500, is better because leading to more accurate results.

Other differences between these varieties stand out to the eye at once:

1. The two varieties differ as to the natural length of ear, because the largest number of ears of any one length (65) in the first variety is ten inches long, while the largest of any one length in the second variety (94) is but nine inches long.

2. This shows a strong tendency in the first variety to assume a ten inch type, but a still stronger tendency in the second variety to assume a nine inch type. We see this tendency to assume a common type is stronger in the second variety than in the first, because a larger proportion of its ears agree in length—94 as compared to 65.

3. We notice that in the second variety this largest frequency is exactly in the middle of the array, the numbers decreasing somewhat uniformly both ways, with three frequencies above and three below.

4. We notice that in the first variety the highest frequency, 65, is considerably above the middle of the array, with three frequencies above and six below.

5. The first array is longer than the second, showing that its type is not only less pronounced but there is a greater tendency to deviate.

The array is thus capable of furnishing at a *glance* a number of interesting differences, but if submitted to well-known mathematical processes it will yield definite values for all these differences. It is the purpose of this circular to explain very briefly the methods of making these computations and securing these values.

THE MODE

The highest frequency in any array, of course, betrays the type which the variety in question tends to assume. That is to say, 65 ears or 32.5 percent of the first variety agree in length (10 inches.) Being larger than any other frequency, it indicates clearly that this variety tends to assume a ten inch type. We, therefore, say that ten is the "mode" of this variety, because it is the length of ear represented by the highest frequency in that variety, and 32.5 percent would be its modal coefficient. For the same reason we say that nine is the mode of the second variety. The value of the mode, therefore, is determined by *inspection* as soon as the array is made, requiring no computation whatever.

THE MEAN

It is quite evident to the observer, however, that the mode does not represent the average or mean length of the ear of the variety. This is evident by the inspection of these two arrays. In the second array the numbers above the modal value, 94, are clearly larger than those below it. That is to say, 43, 18, and 5, are larger than 29, 8, and 3. It is perfectly clear, therefore, that the average length for the second variety is *above* its mode; and by inspection of the first array it will be noted that its average, or mean, is *below* its mode, because the frequencies below 65 are, upon the whole, greater than those above.

In the attempt to compute this mean it is evident that we cannot secure it by simply averaging the two extremes of the array, because the different measurements are represented by different numbers. That is to say, we cannot take the mean between 4 and 13 ($8\frac{1}{2}$) and declare that to be the average length of the ear of the first variety. It would be a true average if all these lengths were equally represented, but not being equally represented, it is necessary as the mathematicians say, to "weight" these measurements. That is to say we should multiply each length of ear by its frequency before we take the average. Therefore we have for the average of the first array the following:

$(4 \times 2) + (5 \times 4) + (6 \times 7) + (7 \times 12) + (8 \times 22) + (9 \times 40) + (10 \times 65) + (11 \times 34) + (12 \times 11) + (13 \times 3) = 1885$, which divided by 200, the total number of ears, gives the number 9.4, which is the true mean or average length of ear of the first variety, so far as it can be accurately represented by 200 ears.

Treating the second array in the same manner we have $(6 \times 3) + (7 \times 8) + (8 \times 29) + (9 \times 94) + (10 \times 43) + (11 \times 18) + (12 \times 5) = 1840$, which divided by 200 the total number of ears, gives the number, 9.2, which is the mean of the average length of ear for the second variety.

It is noticeable that while the mode of the first variety is an inch longer than that of the second, the average length of the ear is but slightly different; and it is possible to imagine two arrays of such a distribution that the one with the higher mode might have the lower mean.

STANDARD DEVIATION

It is perfectly clear that this mean, while an accurate expression of the average length of ear, gives no idea of variability. These two means are very close together, yet the ears differ widely as to distribution, and as we have already noticed the

first variety is much more variable. We must have an expression, therefore, calculated from this mean which will give us some notion of the different spread of these two arrays; or in other words of the range of deviation.

Considering for a moment the first array it will be noticed that the two ears which were four inches long departed or deviated 5.4 inches from the mean of that variety; that the ears which were five inches long deviated 4.4 inches, and so on for the different frequencies. Proceeding upon this basis we write the deviation in each frequency in both arrays as follows:²

Length of Ear.....	4	5	6	7	8	9	10	11	12	13
Frequency, 1st Var.	2	4	7	12	22	40	65	34	11	3
Mean, 1st Var., 9.4.										
Deviations	5.4	4.4	3.4	2.4	1.4	0.4	-0.6	-1.6	-2.6	-3.6

Frequency, 2nd Var.	0	0	3	8	29	94	43	18	5	0
Mean, 2nd Var., 9.2.										
Deviations.....			3.2	2.2	1.2	0.2	-0.8	-1.8	-2.8	

By consulting these deviations we discover that the two ears of the first variety which were four inches long deviated from the mean of that variety 5.4 inches. The four ears that were five inches long deviated from the mean 4.4 inches, and so on.

The seven ears that were six inches long in the first variety deviated 3.4 inches from its mean, but in the second variety the three ears that were six inches long deviated 3.2 inches from *its* mean, and so on for the different frequencies. The problem now becomes one of averaging these deviations and the process agreed upon by mathematicians is as follows:

Square each deviation, multiply by its frequency, divide by the number of ears, and extract the square root. The process carried out in full is as follows:

² The minus signs have no significance here. It is the custom of mathematicians to consider all values to the *right* as negative. It does not necessarily mean "*less*."

First Array.—

$5.4^2 \times 2$	$=$	$5.4 \times$	5.4×2	$=$	58.32
$4.4^2 \times 4$	$=$	$4.4 \times$	4.4×4	$=$	77.44
$3.4^2 \times 7$	$=$	$3.4 \times$	3.4×7	$=$	80.92
$2.4^2 \times 12$	$=$	$2.4 \times$	2.4×12	$=$	69.12
$1.4^2 \times 22$	$=$	$1.4 \times$	1.4×22	$=$	43.12
$0.4^2 \times 40$	$=$	$0.4 \times$	0.4×40	$=$	6.40
$-0.6^2 \times 65$	$=$	$-0.6 \times$	0.6×65	$=$	23.40
$-1.6^2 \times 34$	$=$	$-1.6 \times$	-1.6×34	$=$	87.04
$-2.6^2 \times 11$	$=$	$-2.6 \times$	2.6×11	$=$	74.36
$-3.6^2 \times 3$	$=$	$-3.6 \times$	3.6×3	$=$	38.88
					<u>559.00</u>

Divided by 200 = 2 975. Of which the square root is 1.6 = Standard Deviation.

Second Array.—

$3.2^2 \times 3$	$=$	$3.2 \times$	3.2×3	$=$	30.72
$2.2^2 \times 8$	$=$	$2.2 \times$	2.2×8	$=$	38.72
$1.2^2 \times 29$	$=$	$1.2 \times$	1.2×29	$=$	41.76
$0.2^2 \times 94$	$=$	$0.2 \times$	0.2×94	$=$	3.76
$-0.8^2 \times 43$	$=$	$-0.8 \times$	-0.8×43	$=$	27.52
$-1.8^2 \times 18$	$=$	$-1.8 \times$	-1.8×18	$=$	58.32
$-2.8^2 \times 5$	$=$	$-2.8 \times$	-2.8×5	$=$	39.20
					<u>240.00</u>

Divided by 200 = 1.20. Of which the square root is 1.1 = Standard Deviation.

These two computations give 1.6 + for the first array, and 1.1 for the second array, and are good expressions to indicate the average deviation, or what is commonly called "standard deviation."

It is a good expression for the average variability and may be understood to mean that in the first variety the average tendency to deviate from its mean length is 1.6 inches, while the average tendency of the second variety to depart from its mean length is less, being only 1.1 inches.

COEFFICIENT OF VARIABILITY

If now each standard deviation be divided by its mean we shall have an accurate expression of the variability of its variety. That is to say 1.6 divided by 9.4 equals .17 percent, and 1.1 di-

vided by 9.2 equals .12 percent. These two values may well be taken as fairly representing the variability of the varieties in question.

DEVIATION FROM TYPE

The breeder is often more interested in deviation from type than in deviation from mean. If then he should use the *mode* instead of the *mean* as a base from which to make his calculations, he would get a value that would express the deviation from type. This is often important to know where questions of selection are involved.

These same methods are applicable to all forms of variability that can be accurately measured, weighed, or counted. We may, therefore, in the same way study the circumference and weight of ears, the number of rows, and the percentage of grain on the ear, size or height of stalk in corn, or we may apply these methods to any form of breeding with either plant or animal.

The coefficient of variability, being an abstract term, we can compare the variability of corn with that of any other species or any other variable—as to the length of leg of the horse, the amount he can pull, or the rate at which he can travel. It can also be used regarding the milk produced by cows, the percentage of butter-fat, the amount of food consumed and the period of lactation, the dimensions of the human body, or any other form of variability.

Perhaps few breeders will care to make all the determinations herein suggested; but the array is easily made where sufficient numbers can be found, and the mode and the modal coefficient can be easily determined. Very much can be learned of variability by inspection of this array and by a few exceedingly simple computations, and it is all worth while because success in breeding will depend very largely upon the familiarity of the breeder with the type he is handling and upon his knowledge of inherent tendencies as compared with his own standard of selection.

